EasyChair Preprint

№ 10441

Inference for Variance-Gamma Driven Stochastic Systems

Joseph Johnson, Yaman Kindap and Simon Godsill

June 23, 2023

Inference for Variance-Gamma Driven Stochastic Systems

Joseph Johnson
Department of Engineering
University of Cambridge
Cambridge, UK

Yaman Kındap
Department of Engineering
University of Cambridge
Cambridge, UK

Simon Godsill
Department of Engineering
University of Cambridge
Cambridge, UK

***Abstract*—In this work we present the variance-gamma driven state-space model (VGSSM) - a linear vector stochastic differential equation driven by the variance-gamma (VG) Lévy process, and propose a novel inference framework in such systems. There are closed form expressions for the first four moments of the marginals of the VG process, allowing for more flexible modelling than Brownian motion (BM), retaining BM as a limiting case. The conditionally Gaussian formulation of the variance-gamma process lends itself well to the use of a marginalised particle filter (MPF) which can include the estimation of model parameters as part of the sampling framework. As an example we present a state-space formulation of Langevin dynamics in the VGSSM for estimation of both the observed and the latent first-order dynamics of a system. We apply this specific Langevin formulation to synthetically generated data to validate the results of the MPF, followed by an application to foreign-exchange tick data to demonstrate the method for trend tracking in data sets that are irregularly sampled in time.**

***Index Terms*—non-Gaussian stochastic process, sequential Monte Carlo, nonlinear filtering, particle filter, Lévy process, stochastic differential equation**

I. Introduction

There has been recent interest in modelling dynamic systems subject to uncertainty and randomness through stochastic differential equations (SDEs) such as in [1], [2]. These models provide a continuous-time representation of a randomly evolving system and are applicable to all data patterns, in particular to irregularly sampled data sets. Typically it is assumed that the random noise process can be characterised by a Brownian motion term through Central-limit theorem arguments. However it is well-known that many real world systems exhibit non-symmetric and heavy tailed behaviour that require an extended model for the noise process, such as in financial systems [3], communications [4], signal processing [5], image analysis [6], audio processing [7], [8], and in climatological sciences. In such systems a Lévy process can generalise the stochastic driving noise to include behaviour ranging from Poisson processes to α-stable processes.

In this work we study linear vector stochastic differential equation models [9] driven by a variance-gamma (VG) process [10], [11]

$$d\mathbf{X}(t) = \mathbf{A}\mathbf{X}(t)dt + \mathbf{h}dZ(t) \tag{1}$$

where $Z(t)$ is the Background-Driving Lévy process (BDLP) [12] characterised by a VG process, $\mathbf{X}(t)$ is a state vector containing the system variables at time t; $\mathbf{A}$ and $\mathbf{h}$ are system matrices which define the interactions between states over time and the driving process. Recent work in such models using the α-stable process can be found in [13]–[15]. In contrast with the stable process which has infinite variance and typically undefined mean, the VG process is attractive in practice since the first four moments of the marginal distribution are finite and exist in closed form.

The three parameter VG process is studied in [10], [11] as a time-changed Brownian motion $\mathcal{V}(t) = B(\Gamma(t))$ where $B(t; \mu, \sigma^2)$ is a Brownian motion with drift, scale parameters μ, σ respectively and a *subordinator* $\Gamma(t; \alpha, \beta)$ characterised by a gamma Lévy process with drift, rate parameters α, β. In the case of VG processes the drift rate α is set to 1 which represents a prior assumption on the system dynamics that will be discussed in Section II. Extending our model to include other values of α is straightforward and leads to the more general case of normal-gamma processes.

Many different and useful models can be adapted to the general structure of (1), by defining particular forms for $\mathbf{A}$ and $\mathbf{h}$ in (2), including the continuous-time autoregressive (CAR), the CAR moving-average (CARMA) [16], [17] and the Ornstein-Uhlenbeck (OU) processes [18], as well as all of the standard spatial linear tracking models [1]. As a specific case, consider the following formulation of Langevin dynamics

$$d\dot{X}(t) = \theta\dot{X}(t)dt + dZ(t)$$

that finds application in physics and biological sciences. The rate of reversion to zero is proportional to the magnitude of $\dot{X}(t)$ and the (usually negative) decay parameter θ. We will interpret $\dot{X}(t)$ as a latent trend process for the data, which is typically observed only partially through its integral $X(t)$. The following state-space formulation forms the basis of our model

$$\begin{bmatrix} dX(t) \\ d\dot{X}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & \theta \end{bmatrix} \begin{bmatrix} X(t) \\ \dot{X}(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dZ(t) \tag{2}$$

using the fact that $dX(t) = \dot{X}(t)dt$ by definition.

Here we develop a novel framework for inference of states in the general dynamical model of (1) driven by a VG process, defined as the variance-gamma driven state-space model (VGSSM), and use the Langevin dynamics as a specific case. The new methods are structured similarly to those in

[14] with alterations to deal with the specifics of the VG driving process. The paper is organised as follows. In Section II we review the probabilistic background required for Lévy processes and subordinators and then present gamma processes and VG processes along with their simulation algorithms. In Section III we review some fundamental results on the simulation of vector SDEs. In Section IV we present novel formulations of Langevin dynamics as a Lévy state-space model together with the design of a marginalised particle filter for inference. In Section V we present the application of the Langevin model to synthetic and real datasets, providing an initial demonstration of the new methods. Lastly, in Section VI we review the methods studied in this work and discuss potential extensions to our model and inference algorithm. A code repository for the application of VGSSMs is made available in Python[1].

II. Theoretical Background

In this section we review the series representation of Lévy processes and present simulation algorithms for the gamma and variance-gamma process.

Let $\{Z(t) : t \geqslant 0\}$ be a pure jump Lévy process having no drift or Brownian motion part. According to the Lévy-Khintchine representation, the characteristic function $\phi_Z(s,t)$ can be expressed as ([19], Corollary 13.8)

$$\phi_Z(s,t) := \mathbb{E}\left[\left(\exp\left(isZ(t)\right)\right] = \exp\left(t\psi(s)\right)\right.$$

where the characteristic exponent is

$$\psi(s) := \int_{\mathbb{R}\setminus\{0\}} \left(\exp\left(isx\right) - 1 - isx\mathbb{I}\{|x|<1\}\right)Q(dx)$$

Here $\mathbb{I}\{\cdot\}$ denotes the indicator function and $Q(dx)$ is the Lévy measure satisfying $\int_{\mathbb{R}\setminus\{0\}} \min(1,x^2)Q(dx) < \infty$ which defines the law of the random jump sizes J.

In this work we are also interested in a more constrained class of non-negative, non-decreasing, pure jump Lévy processes $W(t)$ called *subordinator* processes where the characteristic exponent simplifies to

$$\psi_W(s) = \int_0^\infty (e^{isx} - 1)Q_W(dx)$$

with the more restrictive requirement that $\int_0^\infty \min(1,x)Q_W(dx) < \infty$. Subordinators are in general used to measure the level of underlying activity in a system by changing the effective rate of passage of a Brownian motion which result in non-Gaussian heavy-tailed behaviour.

A. The gamma process

The VG process is defined as a subordinated Brownian motion where the subordinator is a gamma process $\Gamma(t;\alpha,\beta)$ with the two parameter Lévy measure

$$Q_\Gamma(dx) = \frac{\alpha^2}{\beta}x^{-1}\exp\left(-\frac{\beta}{\alpha}x\right)dx, \quad x>0$$

[1]https://github.com/jj2249/vglm

and gamma distributed marginals with shape parameter $\frac{\alpha^2 t}{\beta}$, rate parameter $\frac{\alpha}{\beta}$ and density f_Γ such that

$$f_\Gamma(\gamma;t,\alpha,\beta) = \frac{1}{\Gamma(\frac{\alpha^2 t}{\beta})}\left(\frac{\alpha}{\beta}\right)^{\frac{\alpha^2 t}{\beta}}\gamma^{\frac{\alpha^2 t}{\beta}-1}e^{-\frac{\alpha}{\beta}\gamma}, \quad \gamma>0$$

corresponding to a drift rate α and a variance rate β, i.e. $\mathbb{E}\left[f_\Gamma(\gamma;t,\alpha,\beta)\right] = \alpha t$, $\mathbb{V}\text{ar}\left[f_\Gamma(\gamma;t,\alpha,\beta)\right] = \beta t$, where $\mathbb{E}[\cdot]$ is the expectation operator.

Both gamma and VG processes are examples of infinite activity Lévy process such that $\int Q(dx) \to \infty$. The simulation of such processes are in general intractable because they possess an infinite number of *small* jumps in any finite time interval. For infinite activity processes with non-negative increments the inverse Lévy (IL) method provides a way to model the process as an infinite series of decreasing random variables and simulate the underlying jumps by truncating the series after a finite number of terms [20]. The IL method requires the associated tail mass of the Lévy density to be an invertible function, this method is generalised in [21] through the use of rejection sampling methods to include intractable densities, and the random Poisson truncations of these infinite series are studied in [22]. This formulation is known as the generalised shot noise method and will be used throughout this work for the simulation of the driving VG process.

For a Lévy jump process $Z(t)$ with Lévy measure Q, we define the upper tail density as $\Pi(x) := \int_x^\infty Q(dx)$. The IL method [21] utilises the inverse upper tail density $\Pi^{-1}(y) = \inf\{x \in \mathbb{R} : \Pi(x) < y\}$ to write the following series representation for $Z(t)$

$$Z(t) = \sum_{i=1}^{\infty} \Pi^{-1}(E_i)\mathbb{I}\{V_i \leqslant t\} \tag{3}$$

where $\{E_i\}_{i=1}^\infty$ are the epochs of a unit rate Poisson process, straightforwardly generated as a cumulative sum of exponential random variables, and $\{V_i\}_{i=1}^\infty$ are random arrival times of the jumps, uniform on $[0,T]$. The *thinning* or *rejection sampling* method circumvents the need to invert $\Pi(x)$ for these intractable processes by simulating jumps from a *dominating* process with a Lévy measure $Q_0(dx)$ satisfying $\frac{Q(dx)}{Q_0(dx)} \leqslant 1, \quad \forall x \geqslant 0$. Each jump is then accepted with probability $a(x) = \frac{Q(x)}{Q_0(x)}$ (assuming the existence of density functions for each measure).

For gamma processes, a dominating Lévy measure with density $Q_0(dx)$ can be defined as [21]

$$Q_0(dx) = \frac{\alpha^2}{\beta}\left(1+\frac{\alpha}{\beta}x\right)^{-1}dx$$

which has closed form inverse tail density

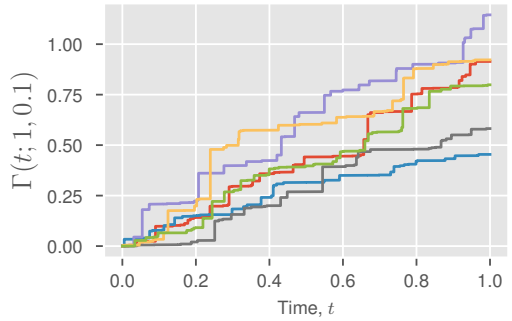
$$\Pi_0^{-1}(y) = \left[\frac{\alpha}{\beta}\left(\exp\left(\frac{\beta}{\alpha^2}y\right)-1\right)\right]^{-1} \tag{4}$$

and acceptance (thinning) probability

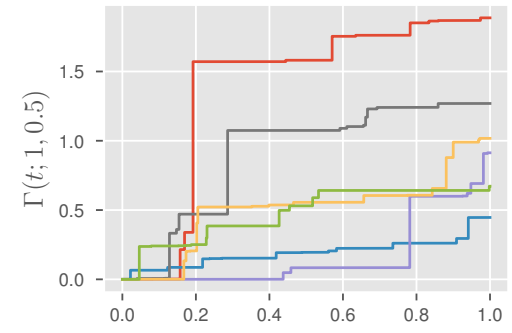
$$a(x) = \left(1+\frac{\alpha}{\beta}x\right)\exp\left(-\frac{\alpha}{\beta}x\right) < 1 \tag{5}$$

Algorithm 1 Simulation of sample paths of a Gamma process $\{\Gamma(t;\alpha,\beta) : 0 \leqslant t \leqslant T\}$.

1) Generate G_s exponential random variables, $\{e_i\}_{i=1}^{G_s}$ with rate parameter $\frac{1}{T}$
2) Calculate the cumulative sum of $\{e_i\}_{i=1}^{G_s}$ to give Poisson epochs $\{E_i\}_{i=1}^{G_s}$
3) Calculate jump sizes $\{d\Gamma_i\}_{i=1}^{G_s} = \{\Pi_0^{-1}(E_i)\}_{i=1}^{G_s}$ according to (4)
4) Calculate acceptance probabilities $\{a_i\}_{i=1}^{G_s} = \{a(d\Gamma_i)\}_{i=1}^{G_s}$ according to (5)
5) For each jump $d\Gamma_i$, accept with probability a_i leaving N_s accepted samples $\{d\tilde{\Gamma}_i\}_{i=1}^{N_s}$
6) For each accepted sample, generate a corresponding jump time $\{V_i\}_{i=1}^{N_s} \overset{\text{iid}}{\sim} \mathcal{U}[0,T]$
7) Return $\{(d\tilde{\Gamma}_i, V_i)\}_{i=1}^{N_s}$,
8) Optionally, form $\Gamma(t;\alpha,\beta) = \sum_{i=1}^{N_s} d\tilde{\Gamma}_i \mathbb{I}\{V_i \leqslant t\}$



(a) $\beta = 0.1$



(b) $\beta = 0.5$

Fig. 1: Samples from the gamma process, $\Gamma(t;1,\beta)$. Six samples each for $\beta = 0.1, 0.5$.

A summary of the simulation procedure is shown in Alg. 1. Fig. 1 shows independent samples from the gamma process with $\alpha = 1$. The process moves by distinct positive jumps arriving at random times with β controlling both the rate of arrival, and the size of the largest jumps. The number of samples required for proper convergence of the processes increases for decreasing β.

B. The variance-gamma process

The variance-gamma process is constructed as a time-changed Brownian motion, or subordinated process, with subordinator $\Gamma(t;1,\beta)$. Setting the drift rate as $\alpha = 1$ can be interpreted as a prior modelling assumption where $\mathbb{E}[\Gamma(t;1,\beta)] = t$. The VG process $\mathcal{V}(t;\mu,\sigma^2,\beta)$ can be expressed in terms of a standard Brownian motion and a gamma process as

$$\mathcal{V}(t;\mu,\sigma^2,\beta) = \mu\Gamma(t;1,\beta) + \sigma B(\Gamma(t;1,\beta);0,1)$$

where the rate parameter β controls excess kurtosis over Brownian motion, σ controls scale and μ controls skewness.

The Lévy measure of the pure jump VG process is defined as

$$Q_{\mathcal{V}}(dx) = \frac{\exp\left(\frac{\mu}{\sigma^2}x\right)}{\beta|x|}\exp\left(\frac{1}{\sigma}\sqrt{\frac{2}{\beta}+\frac{\mu^2}{\sigma^2}}|x|\right)dx, \quad x \in \mathbb{R}$$

The marginal distribution of the VG process, the variance-gamma distribution can be expressed in terms of a Bessel function as [11]

$$f_{\mathcal{V}}(v;t,\mu,\sigma^2,\beta) = \frac{2\exp\left(\frac{\mu v}{\sigma^2}\right)}{\beta^{t/\beta}\sqrt{2\pi\sigma^2}\Gamma\left(\frac{t}{\beta}\right)}\left(\frac{v^2}{\delta}\right)^{\tau/2}K_\tau\left(\frac{|v|}{\sigma^2}\sqrt{\delta}\right)$$

where $v \in \mathbb{R}$, $\tau = \frac{t}{\beta} - \frac{1}{2}$, $\delta = \frac{2\sigma^2}{\beta} + \mu^2$ and K_ν is the modified Bessel function of the second kind with index ν. The first four moments of the VG distribution are known in closed form as [11]

$$\begin{aligned}
\mathbb{E}\left[f_{\mathcal{V}}(v;t,\mu,\sigma^2,\beta)\right] &= \mu t \\
\mathbb{V}\text{ar}\left[f_{\mathcal{V}}(v;t,\mu,\sigma^2,\beta)\right] &= \mu^2\beta t + \sigma^2 t \\
\mathbb{S}\left[f_{\mathcal{V}}(v;t,\mu,\sigma^2,\beta)\right] &= 2\mu^3\beta^2 t + 3\sigma^2\mu\beta t \\
\mathbb{K}\left[f_{\mathcal{V}}(v;t,\mu,\sigma^2,\beta)\right] &= 3\sigma^4\beta t + 12\sigma^2\mu^2\beta^2 t + 6\mu^4\beta^3 t \\
&\quad + 3\sigma^4 t^2 + 6\sigma^2\mu^2\beta t^2 + 3\mu^4\beta^2 t^2
\end{aligned}$$

The process is symmetric for $\mu = 0$ and the sign of μ controls whether the skew is positive or negative. In the unskewed case, the variance and kurtosis expressions simplify to $\sigma^2 t$ and $3\sigma^4 t(\beta + t)$ respectively. The effects of these parameters on the marginal distribution is visualised in Fig. 2. When $\beta \to 0$ we recover Brownian motion since $\Gamma(t) \to t$; thus Brownian motion can be viewed as a special limiting case of the VG process.

The inverse tail density for the VG process does not admit a closed form expression. Instead we use the normal variance-mean mixture representation [23] of the jumps of the VG process[2] where $\{d\mathcal{V}_i\}_{i=1}^{\infty}$ can be directly written in terms of the jumps of a gamma subordinator $\{d\Gamma_i\}_{i=1}^{\infty}$ as

$$d\mathcal{V}_i | d\Gamma_i \sim \mathcal{N}(\mu d\Gamma_i, \sigma^2 d\Gamma_i) \tag{6}$$

Our algorithm for simulation of the VG process is outlined in Alg. 2. Note that the conditionally Gaussian form in (6) is

[2]Here and elsewhere we use a minor abuse of notation in that $\{d\mathcal{V}_i\}$ (resp. $\{d\Gamma_i\}$) denotes the set of *non-zero* increments of $\mathcal{V}(t)$ (resp. $\Gamma(t)$), i.e. its *jumps*.

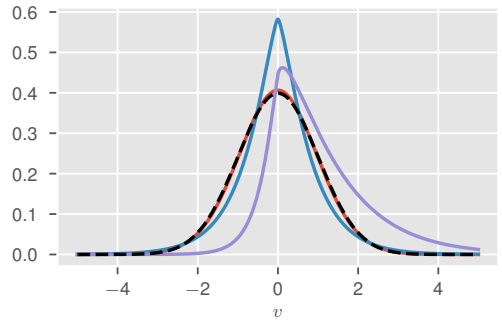


Fig. 2: Marginal distributions of the VG process at time $t = 1$ with scale $\sigma^2 = 1$. Blue: high excess kurtosis, $\beta = 0.75$. Purple: positive skew, $\beta = 0.75$, $\mu = 1$. Red: low excess kurtosis, $\beta = 0.05$ with standard normal dashed-black for comparison.

Algorithm 2 Simulation of sample paths of a VG process $\{\mathcal{V}(t; \mu, \sigma^2, \beta) : 0 \leqslant t \leqslant T\}$.

1) Sample a single set of jump sizes and times from the gamma process, $\{(d\Gamma_i, V_i)\}_{i=1}^{N_s}$, using Alg. 1
2) Sample N_s jumps from the conditional in Eq. (6)
3) Form $\mathcal{V}(t; \mu, \sigma^2, \beta) = \sum_{i=1}^{N_s} d\mathcal{V}_i \mathbb{I}\{V_i \leqslant t\}$

particularly useful for efficient inference algorithms and will be extensively used in this work. Note also that the general form of the conditionally Gaussian jump distribution in (6) is different from the one required for the α-stable Lévy process considered in [14].

Example sample paths from the VG process are shown in Fig. 3 where a fixed number N_s of jumps were generated. Adaptive schemes can be incorporated as discussed in [24].

III. Simulation of vector SDEs

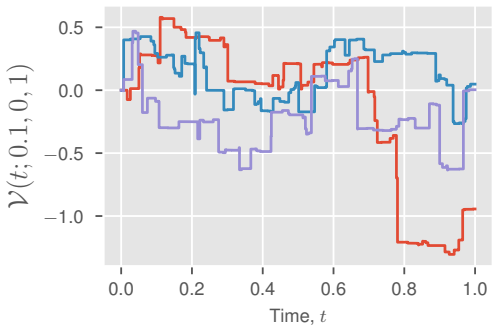
In this section, we review the basics of simulating vector SDEs driven by pure jump Lévy processes. The interested reader can refer to [9] for a more detailed study on SDEs.

The discrete equivalent of (1) will be familiar to many systems engineers, often driven by Gaussian white noise. The noise process $Z(t)$ in our model is a continuous-time stochastic process, therefore the solution for the state at time t is itself a continuous-time stochastic process defined by a stochastic integral and an initial condition $\mathbf{X}(0)$

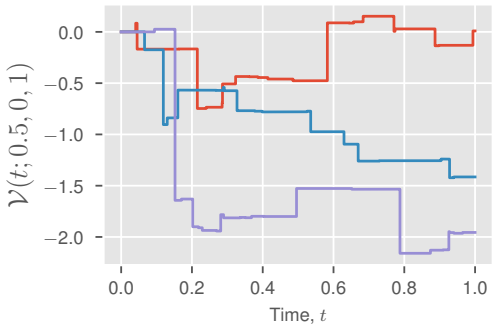
$$\mathbf{X}(t) = e^{\mathbf{A}t}\mathbf{X}(0) + \int_0^t e^{\mathbf{A}(t-u)}\mathbf{h}dZ(u) \tag{7}$$

The first term is deterministic and the second term is the cumulative effect of passing the full history of the noise process through the linear system. Define the following stochastic process

$$\mathbf{I}(\mathbf{f}_t) := \int_0^\infty \mathbf{f}_t(u)dZ(u) \tag{8}$$



(a) $\beta = 0.1$



(b) $\beta = 0.5$

Fig. 3: Samples from the VG process, $\mathcal{V}(t; \beta, 0, 1)$. Three samples each for $\beta = 0.1, 0.5$.

with $\mathbf{f}_t(u) := e^{\mathbf{A}(t-u)}\mathbf{h}\mathbb{I}\{u \leqslant t\}$. Then (7) can be expressed as

$$\mathbf{X}(t) = e^{\mathbf{A}t}\mathbf{X}(0) + \mathbf{I}(\mathbf{f}_t)$$

which is the sum of a deterministic and a stochastic process. This formulation is key to the inference algorithms in Section IV. We propose that when $Z(t)$ is the VG process, $\mathcal{V}(t)$, we can rewrite $\mathbf{I}(\mathbf{f}_t)$ in the following form

$$\mathbf{I}(\mathbf{f}_t) = \sum_{i=1}^{\infty} \mathbf{f}_t(V_i)d\mathcal{V}_i = \sum_{i:V_i \in (0,t]} e^{\mathbf{A}(t-V_i)}\mathbf{h}d\mathcal{V}_i \tag{9}$$

A full proof of this proposal, to be presented in a subsequent publication, follows a similar procedure to Theorem 6.4 in [25] by considering the limit of the point process representation of $\mathcal{V}(t)$ in terms of its jumps $\{(d\mathcal{V}_i, V_i)\}$

$$\mathcal{V}^{(n)}(t) = \sum_{i=1}^{n} d\mathcal{V}_i \mathbb{I}\{V_i \leqslant t\}$$

which converges uniformly on $(0, T]$ to the VG process as $n \to \infty$, i.e.

$$\sum_{i=1}^{n} \mathbf{f}_t(V_i)d\mathcal{V}_i \overset{n\to\infty}{\to} \int_0^\infty \mathbf{f}_t(u)d\mathcal{V}(u)$$

with which we have recovered a discrete sum (albeit countably infinite) representation for the continuous time process $\mathbf{I}(\mathbf{f}_t)$.

In contrast to [14] who substitute the shot-noise representation of the general Lévy process shown in (3) into (9) to find an expression for $\mathbf{I}(\mathbf{f}_t)$ in terms of Poisson epochs, we express the formulae directly in terms of the jumps of the VG process. We hereinafter write the state at time t, $\mathbf{X}(t)$, as $\mathbf{X}_t$ for the sake of brevity.

An iterative formula to simulate a sample path from (1) starts from $\mathbf{X}_0 = \mathbf{0}$ and simulates increments on each interval $(s, t]$ as

$$\mathbf{X}_t = e^{\mathbf{A}(t-s)}\mathbf{X}_s + \sum_{i:V_i \in (s,t]} e^{\mathbf{A}(t-V_i)}\mathbf{h} dZ_i \tag{10}$$

This is natural when modelling observed data sequences with possibly non-uniform arrival times, for which a prediction at the time of the incoming observation can be made with knowledge of $t - s$ and $\mathbf{X}_s$ only. This representation cannot be simulated exactly since we must truncate the small jumps at some finite value in each sub-interval. However, our methods provide a monotonically decreasing set of random jump sizes which ensures that only the smallest jumps less than a threshold of say c are excluded from the resulting series. The accuracy of the approximation $\mathbf{I}^c(\mathbf{f}_t)$ depends on the threshold c and the moments of the residual error from this calculation $R_t^c = \mathbf{I}(\mathbf{f}_t) - \mathbf{I}^c(\mathbf{f}_t)$ may be characterised exactly (see also [14], [22]) and used to generate approximations to the residual as in [24].

IV. The Variance-Gamma Driven State-Space Model (VGSSM)

This section presents the novel formulation of the VG-driven SDE as a Lévy state-space model. In order to design inference algorithms using the model, we first illustrate how to forward simulate from the model by sampling from the conditional distribution of $\mathbf{X}_t$ given $\mathbf{X}_s$, $s < t$.

A. Forward simulation

Simulation of the VGSSM is rendered simple by the linear Gaussian form (6), which states that the conditional $d\mathcal{V}_i | d\Gamma_i$ is a Gaussian random variable, so that $\mathbf{I}(\mathbf{f}_t)$ is also Gaussian. We can sample $\mathbf{I}(\mathbf{f}_t)$ in (9) by evaluating its mean vector $\mathbf{m}$ and covariance matrix $\mathbf{S}$, as follows:

$$\mathbf{I}(\mathbf{f}_t) | \{V_i, d\Gamma_i\}_{i=1}^{\infty} \sim \mathcal{N}(\mathbf{m}, \mathbf{S})$$

where we have used the fact that conditional on observing the jump times, $\mathbf{f}_t(V_i)$ is deterministic so that

$$\mathbf{m} = \mathbb{E}\left[\mathbf{I}(\mathbf{f}_t)\right] = \sum_{i=1}^{\infty} \mathbf{f}_t(V_i)\mu_{V_i} d\Gamma_i \tag{11}$$

$$\mathbf{S} = \mathbb{V}\mathrm{ar}\left[\mathbf{I}(\mathbf{f}_t)\right] = \sum_{i=1}^{\infty} \mathbf{f}_t(V_i)\mathbf{f}_t(V_i)^T \sigma_{V_i}^2 d\Gamma_i \tag{12}$$

which can be broken down into partial sums on each interval $(s, t]$; we write these as $\mathbf{m}_{(s,t]}$, $\mathbf{S}_{(s,t]}$.

Algorithm 3 Sampling from the VGSSM, $\{\mathbf{X}(t; \mu, \sigma^2, \beta, \theta) : 0 \leqslant t \leqslant T\}$.

1) Generate an increasing set of N random observation times, $\{t_i\}_{i=1}^N$, e.g. as exponential arrivals,
2) Initialise $\mathbf{X}_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$, $t_0 = 0$
3) For $i = 1, \ldots, N$,
   - Generate N_s jumps from a Gamma process using Alg. 1 on the interval $(0, t_i - t_{i-1}]$
   - Calculate truncated sums $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{S}}$ for this set of jumps, Eqs. (11, 12)
   - Sample from the bivariate normal distribution, $\mathbf{n} \sim \mathcal{N}(\mu\tilde{\mathbf{m}}, \sigma^2\tilde{\mathbf{S}})$
   - Increment and store $\mathbf{X}_{t_i} = e^{\mathbf{A}(t_i - t_{i-1})}\mathbf{X}_{t_{i-1}} + \mathbf{n}$
4) Return the discrete path $\{\mathbf{X}_{t_i}\}_{i=1}^N$

At this stage we assume that both μ and σ^2 are static parameters, then define $\mathbf{m} := \mu\tilde{\mathbf{m}}$ and $\mathbf{S} := \sigma^2\tilde{\mathbf{S}}$. The full conditional form for (10) becomes

$$\mathbf{X}_t | \mathbf{X}_s, \tilde{\mathbf{m}}_{(s,t]}, \tilde{\mathbf{S}}_{(s,t]} \sim \mathcal{N}(e^{\mathbf{A}(t-s)}\mathbf{X}_s + \mu\tilde{\mathbf{m}}_{(s,t]}, \sigma^2\tilde{\mathbf{S}}_{(s,t]}) \tag{13}$$

Applying this to the Langevin model, we provide $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{S}}$ when $\mathbf{A}$ and $\mathbf{h}$ are as shown in (2). The analytical forms for $e^{\mathbf{A}(t-s)}$, $\mathbf{f}_t(V_i)$ and $\mathbf{f}_t(V_i)\mathbf{f}_t(V_i)^T$, expressed in terms of $f_t(x) = e^{\theta(t-x)}$ are

$$e^{\mathbf{A}(t-s)} = \begin{bmatrix} 1 & \frac{1}{\theta}\left(f_t(s) - 1\right) \\ 0 & f_t(s) \end{bmatrix}$$

$$\mathbf{f}_t(V_i) = \begin{bmatrix} \frac{1}{\theta}(f_t(V_i) - 1) \\ f_t(V_i) \end{bmatrix}$$

$$\mathbf{f}_t(V_i)\mathbf{f}_t(V_i)^T = \begin{bmatrix} \frac{1}{\theta^2}\left(f_t(V_i) - 1\right)^2 & \frac{1}{\theta}f_t(V_i)\left(f_t(V_i) - 1\right) \\ \frac{1}{\theta}f_t(V_i)\left(f_t(V_i) - 1\right) & f_t(V_i)^2 \end{bmatrix}$$

which are subsequently substituted into (10), (11) and (12). The algorithm for generation of sample paths from this state-space process $\mathbf{X}_t = \mathbf{X}(t; \mu, \sigma^2, \beta, \theta)$ is presented in Alg. 3.

B. Filtering in the state-space model

In this section we present our novel inference procedure for VG driven systems. The linear and conditionally Gaussian substructure of the model lends itself well to the marginalised particle filter (MPF) (otherwise known as the Rao-Blackwellised (RB) filter) (see [26]) for inference in the VGSSM. The aim here is to compute the posterior distribution of a latent state x_t given a causal set of observations $y_{1:t}$, $\pi(x_t | y_{1:t})$. A basic non-marginalised particle filter could be used to generate posterior samples by proposing from the prior gamma process and then using the conditional in (13) to propose the states. The particles would then be weighted according to their likelihood under additive white Gaussian noise

$$g(y_t | x_t) \stackrel{d}{=} \mathcal{N}(y_t; x_t, \sigma^2\kappa_v) \tag{14}$$

so that κ_v scales the noise relative to the variance of the process. Our improved MPF follows the general approach

of [14] by extending the state vector to include the skew parameter μ, $\boldsymbol{\alpha}_t = \begin{bmatrix} x_t & \dot{x}_t & \mu \end{bmatrix}^T$, so that (13) becomes

$$\boldsymbol{\alpha}_t \sim \mathcal{N}(\mathcal{A}_{(s,t]}\boldsymbol{\alpha}_s, \sigma^2 \mathcal{D}\tilde{\mathbf{S}}_{(s,t]})$$

where $\mathcal{A}_{(s,t]}$ is the 3×3 matrix

$$\mathcal{A}_{(s,t]} = \begin{bmatrix} e^{\mathbf{A}(t-s)} & \tilde{\mathbf{m}}_{(s,t]} \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}$$

and $\mathcal{D}$ is the 3×2 matrix

$$\mathcal{D} = \begin{bmatrix} \boldsymbol{I}_{2\times 2} \\ \mathbf{0}_{1\times 2} \end{bmatrix}$$

In this extended state-space μ is modelled as a state variable, allowing its distribution to be computed directly as part of the particle filtering scheme. A possible generalisation to time-varying skew in the model will be studied in future work.

The MPF (see [26]), allows for a significant improvement over this basic case. It proposes random paths directly from the underlying gamma process and uses the conditionally Gaussian sub-structure of the state $\boldsymbol{\alpha}_t$ and the Kalman filter to compute *marginal* weights, with no requirement to simulate $\boldsymbol{\alpha}_t$ itself. This is best understood by considering the case where we know the position and size of the jumps in the underlying gamma process (in the case of synthetic data, say): then the remainder of the model could be solved exactly for $\boldsymbol{\alpha}_t$ using the Kalman filter.

Given the full history of the jumps, the Kalman filter starts from prior state $p(\boldsymbol{\alpha}_1) \overset{d}{=} \mathcal{N}(\boldsymbol{\alpha}_1; \mathbf{a}_{1|1}, \mathbf{C}_{1|1})$, then proceeds by recursive prediction and correction of the parameters for each incoming observation. A prediction is made at each step using J_t, then, a correction is made for y_t. The Kalman prediction based on all information available up until time s and the new jumps J_t is

$$p(\boldsymbol{\alpha}_t | J_{1:t}, y_{1:s}) \overset{d}{=} \mathcal{N}(\boldsymbol{\alpha}_t; \mathbf{a}_{t|1:s}, \mathbf{C}_{t|1:s})$$

and the subsequent correction upon observing y_t is

$$p(\boldsymbol{\alpha}_t | J_{1:t}, y_{1:t}) \overset{d}{=} \mathcal{N}(\boldsymbol{\alpha}_t; \mathbf{a}_{t|1:t}, \mathbf{C}_{t|1:t})$$

The means and covariances can be derived by rewriting observation density (14) as

$$y_t = \mathcal{H}\boldsymbol{\alpha}_t + v_t, \quad \text{with} \quad v_t \sim \mathcal{N}(0, \sigma^2\kappa_v)$$

where $\mathcal{H} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$. Then, the Kalman prediction equations from time s to time t are

$$\mathbf{a}_{t|1:s} = \mathcal{A}_{(s,t]}\mathbf{a}_{s|1:s} \tag{15}$$

$$\mathbf{C}_{t|1:s} = \mathcal{A}_{(s,t]}\mathbf{C}_{s|1:s}\mathcal{A}^T_{(s,t]} + \sigma^2\mathcal{D}\tilde{\mathbf{S}}_{(s,t]}\mathcal{D}^T \tag{16}$$

noting that $\mathcal{A}_{(s,t]}$ is implicitly a function of J_t. The correction step is based on the Kalman gain $\mathbf{K}_t$

$$\mathbf{K}_t = \mathbf{C}_{t|1:s}\mathcal{H}^T(\mathcal{H}\mathbf{C}_{t|1:s}\mathcal{H}^T + \sigma^2\kappa_v)^{-1} \tag{17}$$

then the Kalman correction equations are

$$\mathbf{a}_{t|1:t} = \mathbf{a}_{t|1:s} + \mathbf{K}_t(y_t - \mathcal{H}\mathbf{a}_{t|1:s}) \tag{18}$$

$$\mathbf{C}_{t|1:t} = \mathbf{C}_{t|1:s} - \mathbf{K}_t\mathcal{H}\mathbf{C}_{t|1:s} \tag{19}$$

The MPF estimates the jumps by marginalising them from the posterior, $\pi(\boldsymbol{\alpha}_t, J_t | y_{1:t})$ using a set of N_p weighted samples $\{(\omega_t^{(i)}, J_{1:t}^{(i)})\}_{i=1}^{N_p}$ from the gamma process, to give

$$\pi(\boldsymbol{\alpha}_t | y_{1:t}) \approx \sum_{i=1}^{N_p} \omega_t^{(i)} p(\boldsymbol{\alpha}_t | J_{1:t}^{(i)}, y_{1:t}) \tag{20}$$

with $p(\boldsymbol{\alpha}_t | J_{1:t}^{(i)}, y_{1:t})$ coming from each particle's Kalman filter. This limits the problem to principled estimation of the jumps of the gamma subordinator, thereby reducing the total Monte Carlo variance of the resulting estimators [26]. The weight update for the MPF is

$$\omega_t{}^{(i)} \propto \omega_s^{(i)} p(y_t | y_{1:s}, J_{1:t}^{(i)}, \sigma^2), \quad \text{s.t.} \sum_i \omega_t{}^{(i)} = 1$$

where the conditional likelihood $p(y_t | y_{1:s}, J_{1:t}^{(i)}, \sigma^2)$ is

$$p(y_t | y_{1:s}, J_{1:t}^{(i)}, \sigma^2) = \mathcal{N}(y_t; \mathcal{H}\mathbf{a}_{t|1:t}^{(i)}, \mathcal{H}\mathbf{C}_{t|1:t}^{(i)}\mathcal{H}^T + \sigma^2\kappa_v) \tag{21}$$

which can be readily evaluated upon the arrival of each new observation. Weight degeneracy in the particle filter is monitored using estimated sample size and the standard multinomial resampling with uniform post-selection weights is used [26]. Since the jump proposal density does not depend on scale, we follow the work in [14] to implement full conjugate analysis for σ^2. We can marginalise scale from the full likelihood in (21) to remove the dependence of the weight updates on σ^2. A natural choice of prior for σ^2 is the inverse-gamma prior, $p(\sigma^2) \overset{d}{=} \mathcal{IG}(\sigma^2; \rho, \eta)$ with $\rho = \eta = 10^{-5}$ which provides a closed form posterior distribution for σ^2.

(20) is a mixture-of-Gaussians representation of the posterior density, with $\omega_t^{(i)}$ the weight of the ith Gaussian component. Its mean and covariance are thus

$$\mathbf{a}_{t|1:t}^{\text{mix}} = E(\boldsymbol{\alpha}_t | y_{1:t}) \approx \sum_{i=1}^{N_p} \omega_t^{(i)} \mathbf{a}_{t|1:t}^{(i)} \tag{22}$$

$$\mathbf{C}_{t|1:t}^{\text{mix}} \approx \sum_{i=1}^{N_p} \omega_t^{(i)} \left(\mathbf{C}_{t|1:t}^{(i)} + \mathbf{a}_{t|1:t}^{(i)}\mathbf{a}_{t|1:t}^{(i)T}\right) - \mathbf{a}_{t|1:t}^{\text{mix}}\mathbf{a}_{t|1:t}^{\text{mix}\ T} \tag{23}$$

V. Experiments

In this section we demonstrate the proposed inference algorithm on synthetic data, followed by an application to real data from foreign exchange markets. We use the Langevin formulation to estimate both the dynamics of the observed price and the latent trends in the market, which are fundamental to momentum-based algorithmic trading strategies such as in [27].

In our synthetic example the data generating model, as in (2), is driven by a VG process with $\mu = 1$, $\sigma^2 = 1$, $\beta = 0.5$ and a decay parameter $\theta = -2$. The observed positions x are measured under Gaussian noise with variance $\sigma^2\kappa_v$ where $\kappa_v = 10^{-5}$. The Kalman filter priors were initialised using a standard Gaussian $\mathbf{a}_{1|1} = \mathbf{0}_{3\times 1}$, $\mathbf{C}_{1|1} = \mathcal{I}_{3\times 3}$.

We run the MPF using the true data generating process as a proposal (i.e. using the correct β, θ and κ_v), with the

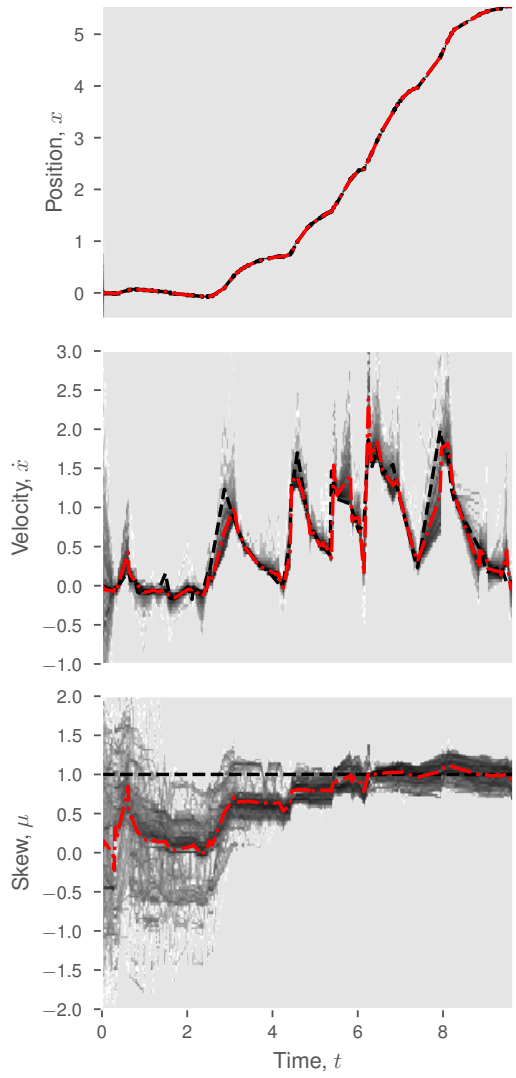


Fig. 4: Filtering example in light observation noise. Black: True, Red: Posterior mean. $\beta = 0.5$, $\mu = 1$, standard normal Gaussian prior, $\kappa_v = 10^{-5}$, $\theta = -2$, 500 particles

results visualised in Fig. 4. The observed positions $\{y_t\}_{t=1}^{100}$ are marked as black dots, black dashed lines represent the underlying latent state of the data generating system and red dot-dashed lines are the inferred mean value of the Gaussian mixture of 500 particles, $\{\mathbf{a}_{t|1:t}^{\text{mix}}\}_{t=1}^{100}$ in (22). The general behaviour of the particle filter is visualised by overlaying these lines on a two-dimensional histogram of the mean trajectories of all particles prior to forming the mixture estimate, $\{\{\mathbf{a}_{t|1:t}^{(i)}\}_{t=1}^{100}\}_{i=1}^{500}$ obtained using the Kalman filter.

The filter is able to track the trend process accurately, including good estimation of the significant jumps - mostly upwards since the skew is positive - and appears to settle on the correct skewness value while maintaining a good diversity of particle paths. We omit the full posterior distribution for σ^2 but present its maximum-a-posteriori estimate $\sigma^2 = 1.055$.

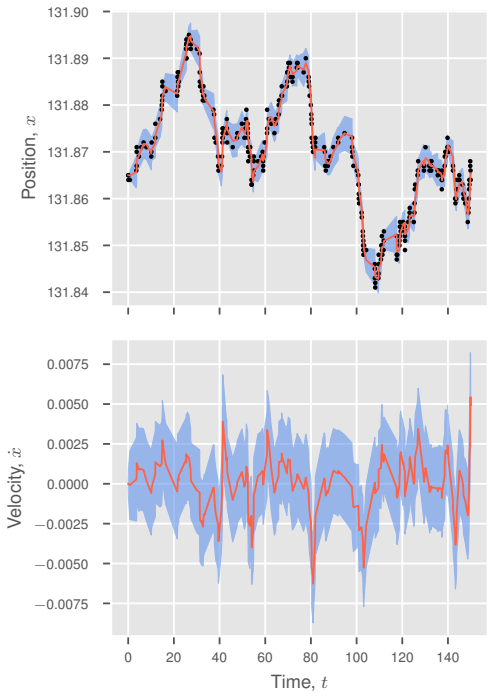


Fig. 5: Test application of the MPF to a real forex CHF-JPY dataset. 300 randomly arriving bid quotes starting at midnight on 1st April, 2022. Black dots are observations, red solid line is the inferred mean and blue band is a 3-standard-deviation interval.

For the real-world data set the parameters of the transition and observation densities, β, θ and κ_v, must be estimated. This is in general a difficult problem, see e.g. [28]. Here we used a two dimensional grid search over the Monte Carlo marginal likelihood of β, θ. κ_v was tuned manually in this case since such parameters are hard to estimate for a model of this complexity. Results are shown in Fig. 5. The upper panel shows a series of randomly arriving bid offers on the Swiss Franc/Japanese Yen foreign exchange for a 150 second period starting at midnight on the 1st April 2022 with a total of 300 observations, sourced online from TrueFX. The estimated parameter values are found as $\beta = 1.003$, $\theta = -0.2778$, $\kappa_v = 0.5$, $\mu = 10^{-6}$ and $\sigma^2_{\text{MAP}} = 1.078 \times 10^{-6}$. The estimated value of β ($\gg 0$) suggests that a Brownian-driven model would be inappropriate compared with the variance-gamma model for this task, owing to the excess kurtosis in the driving stochastic trend process. Simulations we carried out with the corresponding Brownian-driven model demonstrated that in fact this was the case, with large changes in trend not being accurately tracked by the Brownian model.

Fig. 5 shows the inferred filtering distributions for the bid price and the trend using the mean mixture trajectory, along with 3-standard-deviation intervals based on the mixture variances. Jumps in the inferred trend (lower panel) are able to model sudden shifts in the underlying trend of the bid price,

for example sudden persistent downwards shifts around t=80s and 100s. These are likely to be of significant assistance in momentum-based prediction of the data. The inferred skew is omitted as it remains very close to zero throughout, indicating a largely symmetric jump distribution process for this data and time period.

VI. Discussion

The models introduced in this paper provide an expressive representation of continuous-time stochastic linear systems with non-Gaussian properties. Our presentation allows for efficient inference procedures as discussed in Section IV, with the results of our marginalised particle filter (MPF) demonstrated in Section V. Non-Gaussianity in the observed data is handled through the driving VG process which accounts for fluctuations in the underlying activity level of the system. We obtain control over the first four moments of the driving process, which allows modelling of a wide range of real-world phenomena with applications in a variety of fields.

It is worth noting that our formulation of the MPF is valid for any linear vector SDE (1) with well-defined system matrices $\mathbf{A}$ and $\mathbf{h}$. As mentioned earlier, the CAR, CARMA and OU processes can be expressed within the Lévy state-space model formulation, with the forms of the required matrix and vector calculations following the same general scheme as for the Langevin model; in addition the models are readily extended to spatial tracking scenarios in 2 or 3 dimensions. We will explore these possibilities in future application work.

Additionaly our formulation of the VGSSM can be readily extended to include other driving Lévy processes terms possessing the normal variance-mean mixture representation. In particular the generalised hyperbolic (GH) process, which includes the VG process as a special case, is a natural extension of the framework presented here (simulation algorithms for the GIG process, which is the subordinator associated with the GH process, can be found in [29]). The results for SDEs driven by these more general GH processes and more extensive evaluations of the schemes proposed here will be presented in future publications.

References

[1] R. Gan, B. I. Ahmad, and S. J. Godsill, "Lévy state-space models for tracking and intent prediction of highly maneuverable objects," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 4, 2021.

[2] Q. Li, R. Gan, J. Liang, and S. J. Godsill, "An Adaptive and Scalable Multi-Object Tracker Based on the Non-Homogeneous Poisson Process," *IEEE Transactions on Signal Processing*, vol. 71, pp. 105–120, 2023.

[3] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, 2003.

[4] M. de Freitas, M. Egan, L. Clavier, A. Goupil, G. Peters, and N. Azzaoui, "Capacity Bounds for Additive Symmetric α-Stable Noise Channels," *IEEE Transactions on Information Theory*, vol. 63, pp. 5115–5123, 2017.

[5] C. L. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications*, ser. Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley, 1995.

[6] A. Achim, E. E. Kuruoğlu and J. Zerubia, "SAR image filtering based on the heavy-tailed Rayleigh model," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2686–2693, 2006.

[7] S. Godsill and P. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 4, pp. 352–372, 1998.

[8] M. J. Lombardi and S. J. Godsill, "On-line Bayesian estimation of signals in symmetric α-stable noise," *Signal Processing, IEEE Transactions on*, vol. 54, no. 2, pp. 775–779, 2006.

[9] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications (Universitext)*, 6th ed. Springer, Jan. 2014.

[10] D. B. Madan and E. Seneta, "The variance gamma (vg) model for share market returns," *Journal of business*, pp. 511–524, 1990.

[11] D. B. Madan, P. P. Carr, and E. C. Chang, "The variance gamma process and option pricing," *Review of Finance*, vol. 2, no. 1, pp. 79–105, 1998.

[12] O. Barndorff-Nielsen and C. Halgreen, "Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 38, pp. 309–311, 1977.

[13] T. Lemke, M. Riabiz, and S. J. Godsill, "Fully Bayesian inference for α-stable distributions using a Poisson series representation," *Digital Signal Processing*, vol. 47, pp. 96 – 115, 2015.

[14] S. Godsill, M. Riabiz, and I. Kontoyiannis, "The Lévy state space model," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 487–494.

[15] M. Riabiz, T. Ardeshiri and S. Godsill, "Convergence results for tractable inference in α-stable stochastic processes," in *2017 22nd International Conference on Digital Signal Processing (DSP)*, Aug 2017, pp. 1–5.

[16] S. Godsill and G. Yang, "Bayesian inference for continuous-time ARMA models driven by non-Gaussian Lévy processes," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 5, 2006, pp. V–V.

[17] P. Brockwell, "Lévy-driven carma processes," *Annals of the Institute of Statistical Mathematics*, vol. 53, pp. 113–124, 02 2001.

[18] O. E. Barndorff-Nielsen and N. Shephard, "Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 167–241, 2001. [Online]. Available: http://www.jstor.org/stable/2680596

[19] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed. Springer-Verlag, 2002.

[20] T. S. Ferguson and M. J. Klass, "A Representation of Independent Increment Processes without Gaussian Components," *The Annals of Mathematical Statistics*, vol. 43, no. 5, pp. 1634 – 1643, 1972. [Online]. Available: https://doi.org/10.1214/aoms/1177692395

[21] J. Rosiński, "Series representations of Lévy processes from the perspective of point processes," in *Lévy Processes*, O. Barndorff-Nielsen, S. Resnick, and T. Mikosch, Eds. Birkhauser Boston, 2001, pp. 401–415.

[22] S. Asmussen and J. Rosiński, "Approximations of small jumps of Lévy processes with a view towards simulation," *Journal of Applied Probability*, vol. 38, no. 2, pp. 482–493, 2001.

[23] E. Eberlein, "Application of generalized hyperbolic Lévy motions to finance," in *Lévy Processes*, O. Barndorff-Nielsen, S. Resnick, and T. Mikosch, Eds. Birkhäuser, Boston, MA, 2001.

[24] Yaman Kindap and Simon Godsill, "Point process simulation of generalised hyperbolic Lévy processes," 2023.

[25] J. Rosiński, "Tempering stable processes," *Stochastic Processes and their Applications*, vol. 117, no. 6, pp. 677–707, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030441490600144X

[26] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.

[27] H. L. Christensen, J. Murphy, and S. J. Godsill, "Forecasting High-Frequency Futures Returns Using Online Langevin Dynamics," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 4, pp. 366–380, 2012.

[28] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[29] S. Godsill and Y. Kındap, "Point process simulation of generalised inverse Gaussian processes and estimation of the Jaeger integral," *Statistics and Computing*, vol. 32, no. 1, p. 13, Dec 2021. [Online]. Available: https://doi.org/10.1007/s11222-021-10072-0